Edited by Péter Kovács. University of Szeged, January 2018
ISBN 978-963-306-575-4
Proceedings of Challenges and Innovations in Statistics Education Multiplier Conference of ProCivicStat

USING VISUALIZATION TOOLS IN CLASSES

Péter Kovács, Klára Kazár, Éva Kuruczleki
University of Szeged
kovacs.peter@eco.u-szeged.hu
kazar.klara@eco.u-szeged.hu
kuruczleki.eva@eco.u-szeged.hu

Not only statistical method's usage but also their applications on real data have key importance in statistics classes, which can help to understand social- and economic phenomenon. IT tools, real data and visualization tools can mean the link among the statistical methods (for instance correlation, descriptive statistics, regression models) and application fields (for instance ageing society, poverty, income inequalities). Besides applications, statistical literacy should be also improved in the classes.

The mail goal of the paper is to show statistical course materials which are using real statistical data and newer interactive visualization tools (for instance Gapminder or tools for visualizing social and economic networks). The ProCivicStat project is the framework of our work.

INTRODUCTION

The implication of interesting and lifelike topics is highly important in statistics classes. Students should discover the connection between data and real life, and teachers should make the students part of the data set; in this way, their engagement and activity could be increased (Brown, 2016). This paper offers two examples from real life: first, as economic and social networks are deeply embedded in our everyday life, network analysis could be an interesting topic in statistics classes. Second, understanding a social phenomenon such as poverty has a key importance for students in the field of education, hence examining poverty and income inequalities could also be an important topic in statistics classes.

Besides the above mentioned current topics, innovative solutions should also be included in statistics education. University students are the members of Generation Y, who share common characteristics. They are considered the most technologically literate generation, they cannot imagine life without IT tools or Internet. Generation Y cannot be motivated/encouraged by the traditional way of teaching anymore, but they are eager to use technology during classes instead (Eckleberry-Hunt & Tucciarone 2004, Reilly, 2012, Sox, et al. 2014). Using technology, IT and visualization tools in statistics classes provide an appropriate solution for teaching the members of this generation (Chance, et al. 2007, Ridgeway, 2016). More specifically, the usage of personal response systems (PRS or Electronic Voting Systems) is also recommended; they can motivate students to answer questions and they can also confirm whether students have understood the topic covered in the lesson (Lancaster & Titman, 2014).

The ProCivicStat project funded by the European Commission is a strategic partnership among six universities which create resources (theoretical framework, datasets, visualization tools and teaching materials) for statistics education (Engel 2017). One of the goals of this project is to offer citizens topics in statistics education which are relevant for their future life (e.g. economic networks or financial literacy); and another important part of the project is the application of modern solutions (e.g. visualization tools or evaluation, voting systems) in classes to grab the attention of our students. Based on these, the goal of this paper is to introduce two specific teaching methods developed in the framework of ProCivicStat in the topic of network analysis and in the topic of poverty and income inequalities with the help of modern IT solutions.

APPLICATION OF SOCIAL AND BUSINESS NETWORKS

Network studies is an emerging research topic these days as academics in more and more research areas start to realize how deeply embedded networks are in every field of natural and social sciences and even in economics and how well the tools provided by graph theory can be applied to visualize and examine a set of variables describing any phenomena (Barabási, 2016, Roverato, 2017).

Visualizing networks can help the recipient students -either on lower or higher level of education- to better understand the structure of any social or economic phenomena, and to help them discovering the underlying relations between the constituents of a network. Network analysis and graph theory can be introduced in statistical education through various methods. For instance, one way is through showing an adjacency matrix containing correlation coefficients or interdependencies of different variables of a problem, and translating this adjacency matrix into a graph that better visualizes these coefficients, as graphs are great tools for visualizing interdependencies and relations between variables. This way, our aim of enhancing complex thinking of students to better understand complex data and systems can be achieved.

As part of the ProCivicStat project, introductory and advanced level lesson plans are developed. Introductory level lesson plans teach students the basics of graph theory. The lessons start at the very beginning explaining how networks are embedded in our everyday life, just to mention a few examples, transportation network, the internet or even our brain is a network. A network of metro stations, webpages or neurons. Introducing networks through such examples help students to realize that network theory itself is not as distant as they think, and these examples also help them understand faster the notion of nodes, edges, the weight of edges and other properties of a network. After becoming familiar with the building blocks of networks, students can get an insight into creating a graph. We chose Gephi as the main visualization tool in networks and graphs lesson plans, as it is a software solution that contains all the essential features that we need, but simple enough to be quickly understood by the students. Apart from that, Gephi is an open source program that can be downloaded and used entirely for free by any individuals and institutions. Of course, numerous other network analysis and visualization tools are available, our choice is just a guideline, the task lists and lesson plans can be adjusted to any tool the teacher is familiar with or is willing to use. Tasks aiming at the use of visualization tools teach the students how to create simple graphs by first providing them with an input dataset with just a few details –e.g. nodes, edges between them, weight of the edges and whether the edges are directed or not- and then showing them the main steps on how to create the graph itself. Introductory lectures therefore contain exercises that focus mainly on discovering the details of already created graphs and only include basic tasks on creating graphs.

Advanced level lesson plans are when statistics meet graph theory and network science. One might expect that after completing the introductory level lessons, students can instantly continue with the advanced level. That is partially true for the network analysis part, however even though introductory level can be easily understood by even high school students, the statistical part of these lesson plans have higher prerequisites to be obtained previously. These prerequisites include familiarity with basic statistical terms or basic notions of causality and as well some methods of statistical modeling, such as calculating correlation coefficients or applying regression models. Tasks at advanced level begin with a short overview of the main properties of graphs through a visual illustration of a much more complex graph. One example for this task is a graph illustrating correlation coefficients of S&P 500 stock prices through a period of a few years' time: such graph can visualize which stocks moved together in the examined period and highlight to what degree the volatility of one stock can influence others. Apart from learning how to illustrate more complex systems and how to customize it, students can also learn how to cluster nodes of a system with the help of modularity classes, as a result of which they become able to discover certain sub-systems within a network composed of agents that have stronger relations between each other and weaker with other members of the initial system. By completing the advanced level tasks, students obtain the ability to join statistics and network science to

create graphs that provide a deeper insight to a certain phenomenon and learn advanced customization of graphs.

After assessing network studies lesson plans, one might pose the question: why should network studies be included in civic statistics and what benefits students can infer from learning such high-level network studies? The answer is that it is not a must, but an opportunity to include these materials. The main aim of teaching network studies is to develop and enhance complex thinking of recipients and to teach them how to better understand complex systems. It is true that advanced level lesson plans require such statistical knowledge that some target groups might not have, but we felt that it is important to develop advanced level too, as some recipients might want to implement that knowledge in their own research field. It is certainly useful to learn the introductory levels, as even from that lectures, people can realize how deeply embedded networks are in our life and how many phenomena can be understood as a system of nodes connected by edges, and if one might feel that they want to get a better understanding of networks, that is when advanced level tasks can help, in an easy to understand and entertaining way. Network studies therefore even though do not constitute an essential part of civic statistics, certainly provide an additional visualization method and a way of thinking about social and economic phenomena.

APPLICATION FOR EXAMINING POVERTY AND INCOME INEQUALITIES

The aim of this lesson plan is to understand poverty, income inequality and statistical concepts (e.g. correlation) better with the help of Gapminder. Why is it important to create a lesson plan for those topics? First, the examination of poverty and income inequality is important, due to their impacts on our lives. Poverty and income inequality can lead to a widening gap between the rich and the poor, may reduce people's opportunities to study and can have negative effects on people's health (Keeley 2015). Therefore, those topics are considered as actual questions from the point of view of education. Second, Gapminder is a very spectacular dynamic visualization tool, so it is easy to find delightful and stimulating tasks for students. It is possible to download data and to reach the definitions of the indicators. There are some drawbacks: the number of available indicators in the topic of poverty and income inequality is lower compared to Eurostat Databases or to OECD Stat, and there is no option for uploading data to Gapminder. But still, it is important to use Gapminder, because this visualization tool fulfils the needs of Generations Y students.

The target group of this lesson could be university students; because some previous knowledge is needed. Statistical knowledge about concept of correlation coefficient and the logarithmic-linear transformation are needed; and some context knowledge about poverty and income inequalities is necessary. Concerning the structure, the lesson plan consists of two parts. The first part is an introduction of the usage of Gapminder, and in second part students can explore Gapminder themselves as a kind of independent work.

The first part of the lesson focuses on the usage of Gapminder. The lecturer shows

- built-in visualizations in the topic of poverty and income inequality (e.g. Yes, most billionaires live in the US or Wealth & Health of Nations), which may rise students' interest in the Gapminder and can be an enjoyable part in the lesson. At this point, the results seen in visualization (e.g. Which countries are rich? Where is higher income inequality?) are discussed.
- options of visualization (axes/indicators, play button, speed of visualization, chart/map view, country selection, other options)
- data in Gapminder World, where indicators can be searched by topic and the definitions of indicators are also available
- creation of a not built-in visualization e.g.: examining the relationship between poverty (ratio of people below 2$ a day) and food supply.

The second part of the class includes independent work in teams. Each team receives a question or should raise a question themselves, e.g.: What kind of indicators can be connected to poverty (ratio of people below 2$ a day) in Africa? What kind of indicators can be connected to income inequality index (Gini) in South-America? What kind of indicators can be connected to the income share of the poorest 10% of the population? How did poverty (ratio of people below 2$ a day) change in Europe? The questions have to be answered with the help of Gapminder. Finally, the teams present their results at the end of the lesson, when the other teams and the instructor evaluate and give feedback about the work of each team.

As an assessment we can conclude that this lesson plan meet the needs of Generation Y (technology orientation, teamwork, feedbacks, entertainment, openness for visualization tools). At the same time the lesson provides to the students an introduction into the main concepts of poverty and income inequality and into the usage of Gapminder too. However, it could be important for students being able to create independent analysis by using other data which are not found in Gapminder or by using other statistical methods which are not built in Gapminder. The usage of other statistical database or methods could be a next level in the topic of poverty and income inequality education, but this lesson plan offers a basis for the education of university students.

CONCLUSION

Using visual tools in education certainly has the benefit of maintaining the attention of the audience. Apart from that, in our previously introduced lesson plans and teaching materials visuals also serve the purpose of making understanding complex social and economic phenomena easier.

Network studies might seem to be a distant topic for students, however by highlighting the resemblance of everyday structures, such as the transportation system, to a network, the ability of students to understand complex systems can be enhanced greatly. Visual representation of networks helps students to discover connections between the constituents and to explore sub-systems, i.e. clusters, within a bigger network, acquiring skills that can be implemented in other research fields as well. While introductory materials aim more at developing general knowledge of the students of all ages, advanced level network materials are more suitable for researchers or university students.

University students are the target group of developed materials in the topics of poverty and income inequality visualized with the help of Gapminder. Students can learn about serious issues in an entertaining way with the help of visualization tools of Gapminder, while they also develop their digital competences. Gapminder seems to have some limitations regarding the available data and statistical tools, however lessons with Gapminder provide a strong base for further statistical education of university students.

To conclude, in the above chapters we introduced two types of our developed teaching materials within the framework of the ProCivicStat project. Both topics aim at teaching students about given issues, while at the same time improving their general understanding of the world. Software solutions of both types of lesson plans have their limitations, however they seem to be a good starting base for further education in civic statistics.

ACKNOWLEDGMENT

The work reported in this paper was supported in part by ProCivicStat project, a strategic partnership of the Universities of Durham, Haifa, Ludwigsburg, Paderborn, Porto and Szeged, funded by the ERASMUS+ program of the European Commission. However the views and opinions expressed in this paper are those of the authors and do not necessarily reflect those of the funding agency.

REFERENCES

Barabási, A. L. (2016). *Network Science.* Northeastern University, Boston.

Brown, M. (2016). Engaging Students in Quantitative Methods: Real Questions, Real Data. In Engel, J. (ed.): *Promoting understanding of statistics about society. Proceedings of the Roundtable Conference of the International Association of Statistics Education (IASE)*, Berlin: ISI/IASE.

Chance, B., Ben-Zvi, D., Garfield, J., & Medina, E. (2007). The Role of Technology in Improving Student Learning of Statistics. Technology Innovations in Statistics Education, 1(1) Article 2.

Eckleberry-Hunt, J. & Tucciarone, J. (2011). The Challenges and Opportunities of Teaching "Generation Y". *Journal of Graduate Medical Education, 3*(4), 458-461.

Engel, J. (2017). Statistical Literacy for Active Citizenship: A Call for Data Science Education. *Statistical Education Research Journal, 16*(1), 44-49.

Keeley, B. (2015). *Income Inequality: The Gap between Rich and Poor*. Paris: OECD Insights, OECD Publishing.

Lancaster, G. A. & Titman, A. C. (2014). Personal Response Systems as a Learning Aid in an Epidemology Course for Postgraduate Statistics Students. In Makar, K., de Sousa, B. & Gould, R. (eds.): *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics*, Flagstaff: ISI/IASE.

Reilly, P. (2012). Understanding and Teaching Generation Y, *English Teaching Forum, 50*(1), 2-11.

Ridgway, J (2016). Implications of the Data Revolution for Statistics Education, *International Statistical Review, 84*(3), 528-549.

Roverato, A. (2017). *Graphical Models for Categorical Data*. Cambridge University Press.

Sox, C. B., Kline, S. F. & Crews, T. B. (2014). Identifying best practices, opportunities and barriers in meeting planning for Generation Y, *International Journal of Hospitality Management, 36*(January), 244-254